ORIGINAL ARTICLE

WILEY

Content overlap analysis of 64 (hypo)mania symptoms among seven common rating scales

Adrian A. Chrobak[1] | Marcin Siwek[2] | Dominika Dudek[3] | Janusz K. Rybakowski[4]

[1] Department of Psychiatry, Jagiellonian University Medical College, Kraków, Poland

[2] Department of Affective Disorders, Chair of Psychiatry, Jagiellonian University Medical College, Kraków, Poland

[3] Department of Adult Psychiatry, Chair of Psychiatry, Jagiellonian University Medical College, Kraków, Poland

[4] Department of Adult Psychiatry, Poznan University of Medical Sciences, Poznań, Poland

Correspondence
Adrian Andrzej Chrobak, Department of Psychiatry, Jagiellonian University Medical College, Kopernika St. 21A, Kraków 31-501, Poland.
Email: adrian.andrzej.chrobak@gmail.com

Funding information
Foundation for Polish Science (FNP)

Abstract

Objectives: Fried (2017) quantified the overlap of items among seven widely used depression rating scales. The analysis revealed substantial heterogeneity of the depressive syndromes and a low overlap among the scales. To our best knowledge, there are no studies evaluating the content overlap of (hypo)mania scales. The goal of our study, therefore, is to quantify the overlap of items among seven widely used (hypo)manic scales, implementing the methodology developed by Fried (2017).

Methods: Seven commonly used (hypo)manic scales underwent content analysis. Symptom overlap was evaluated with the use of the Jaccard index (0 = *no overlap*, 1 = *full overlap*). In case of every scale, rates of idiosyncratic symptoms and rates of specific versus compound symptoms were calculated.

Results: The seven scales gathered 64 hypo(manic) symptoms. The mean overlap among all of the instruments was low (0.35), the mean overlap of each scale with all others ranged from 0.29% to 0.48%, and the overlap among individual scales ranged from 0.20% to 0.65. Thirty-six percent of symptoms appeared only on one scale. Only 6% of the symptoms appeared on all of the instruments.

Conclusions: We have shown that using (hypo)manic scales as interchangeable measurements may be problematic.

KEYWORDS

bipolar disorder, hypomania, mania, rating scales, symptom overlap

1 | INTRODUCTION

Bipolar disorder (BD) is a severe mental illness characterized by the occurrence of episodes of mania or hypomania, depression, and mixed states (Angst, Ajdacic-Gross, & Rössler, 2015). According to the constellation of those episodes, BD is subclassified as BD I and BD II. The diagnosis of BD I requires only a history or presence of mania. BD II is characterized by at least one hypomanic episode and one major depressive episode (Vieta et al., 2018). The term bipolar spectrum can be conceptualized in two ways: first, as the entirety of all bipolar disorders (BD I, BD II, and BD others; Akiskal, 2007). The second meaning of bipolar spectrum would be a disorder showing the features of bipolarity not meeting the criteria for BD II and fulfilling diagnostic space between BD II and unipolar depression as proposed

by Ghaemi, Ko, and Goodwin (2001). This is also called subthreshold bipolarity (Angst, Cui, & Swendsen, 2010). An 11-nation survey has shown that 2.4% of people have had the diagnosis of BD at some point in their lifetime (Merikangas et al., 2011), while another study suggests that this occurrence may reach even 5.1% (Judd & Akiskal, 2003). For patients presenting for treatment of depression, BD patients sometimes remain underrecognized, and the time lag between initial treatment seeking and the correct diagnosis often exceeds 10 years (Hirschfeld et al., 2000; Rucci, Calugi, Miniati, & Fagiolini, 2013).

Hypomanic episodes separate BD II from unipolar depression. Many BD II patients consider their hypomanic periods to be normal phases of especially productive activity, and a plethora of hypomanic symptoms is perceived by patients as signs of well-being. Given the

fact that those episodes may not cause functional impairment and are limited in their time and severity can make the diagnosis a challenge (Vieta, 2010). The consequence of unrecognized hypomanic episodes is an inadequate treatment leading to the worsening of the course of BD (Antosik-Wójcińska, Stefanowski, & Święcicki, 2015) or increased risk of substance misuse (Klimkiewicz, Klimkiewicz, Jakubczyk, Kieres-Salomoński, & Wojnar, 2015a, 2015b) and suicidal behavior (Rajewska-Rager, Sibilski, & Lepczyńska, 2015; Siwek et al., 2009). This indicates that better detection of bipolar traits, such as hypomanic symptoms in patients previously diagnosed with unipolar depression, is strongly needed. This challenge may be resolved through the application of clinical instruments for the assessment of hypomanic symptoms (Łojko et al., 2016; Siwek et al., 2009). Noteworthy, a growing number of studies are using such scales to measure bipolarity symptoms in clinical (Siwek, Dudek, Jaeschke, et al., 2015) and nonclinical populations (Dudek et al., 2015; Siwek, Dudek, Drozdowicz, et al., 2015). Manic episodes are more severe and life-impairing than hypomanic ones. However, because most symptoms of mania and hypomania differ only in quantitative criteria (their severity and duration) and not the qualitative criteria (Vieta et al., 2018), in the further part of the manuscript, we will use the term of (hypo)manic scales and symptoms. This qualitative overlap between manic and hypomanic symptoms is reflected in the fact that instruments for the screening of past hypomanic symptoms (like Mood Disorder Questionnaire—MDQ, or Hypomanic Checklist 32–HCL-32) also detect lifetime episode of mania (Carvalho et al., 2015). Conversely, most common clinical scales for measuring mania symptoms, like the Young Mania Rating Scale (YMRS) or Bech-Rafaelsen Mania Rating Scale (BRMRS), have cut-off points for hypomanic episodes (Bech, 2002; Young, Biggs, Ziegler, & Meyer, 1979).

Fried, 2017 pointed out that there is a common implied assumption that various instruments measuring the same nosological entity can be used as interchangeable measurements. In the case of depression scales, he has shown that this assumption is wrong and may pose a major challenge to the replicability and generalizability of depression research. The author had performed the content analysis of seven common depression scales. Forty percent of the symptoms appeared in only a single scale, whereas only 12% appeared in all of the instruments. It has been shown that the mean overlap of extracted symptoms was low, and the scales differed in regard to their rates of idiosyncratic symptoms (0–33%). To our best knowledge, there are no such studies evaluating the content overlap of (hypo)mania scales. Given the growing number of studies assessing bipolarity in clinical and nonclinical populations, it is important to evaluate to what degree instruments may serve as interchangeable measures. Being inspired by the work of Fried (2017) and his methodology, we decided to quantify the overlap of items among various widely used (hypo)manic scales.

2 | METHODS

According to Fried (2017), we decided to choose seven common (hypo)manic rating scales. Instruments were chosen based on their citation count, frequency in the PubMed database, appearance in studies comparing multiple scales (Lee et al., 2013; Parker et al.,

2012; Parker, Graham, Rees, Futeran, & Friend, 2012; Parker, Graham, Synnott, & Anderson, 2014; Rybakowski et al., 2012), and inclusion in recent reviews and analyses (Carvalho et al., 2015; Miller, Johnson, & Eisner, 2009; Rucci et al., 2013; Vieta, 2010). We have excluded tools assessing temperamental, character, and personality traits such as temperament evaluation of Memphis, Pisa, Paris, and San Diego, Affective Temperament Questionnaire, or Hypomanic Personality Scale. Initial search based on frequency in PubMed database and citation count resulted in the following scales: YMRS, MDQ, Clinician-Administered Rating Scale for Mania (CARS-M), BRMRS, HCL-32, Altman self-rating mania scale, Internal state scale, Self-report version of The Structured Clinical Interview for Mood Spectrum (MOODS-SR) and Bipolar Spectrum Disorder Scale (BSDS), Mood Swings Questionnaire (MSQ). We have decided to exclude the Internal state scale and MOODS-SR because both instruments present a broader range of items discriminating mood states in BD other than (hypo)mania (Bauer, Vojta, Kinosian, Altshuler, & Glick, 2000; Dell'Osso et al., 2002). Thus, in our opinion, the addition of these tools to the analysis will bias the results through the increase of idiosyncratic items and the decrease of overlap in the measured symptoms. Due to the rapidly increasing number of studies evaluating the MSQ in recent years (Parker & Fletcher, 2013; Parker, Fletcher, Blanch, & Greenfield, 2012; Parker, Fletcher, McCraw, & Hong, 2014; Parker & Graham, 2015; Parker, Graham, et al., 2014; Parker, Graham, Hadzi-Pavlovic, et al., 2012; Parker, Graham, Rees, et al., 2012) and, in our opinion, the brevity and the limited role of Altman self-rating mania scale as the independent measure of (hypo)manic symptoms, we decided to include the former to the analysis.

Finally, seven (hypo)manic rating scales were selected: 32-item HCL-32, 11-item YMRS, 13-item MDQ, 19-item BSDS, 11-item BRMRS, 15-item CARS-M, and 27-item MSQ.

2.1 | Content analysis

We performed content analysis according to the methodology of Fried (2017) to determine the content overlap among the scales. First, we summed the number of all of the items from the seven selected rating scales, receiving a total of 128. Then we screened the scales for the presence of similarly worded items within the instruments in order to combine them so as to avoid inflating the number of symptoms (e.g., Fried, 2017 combined "apparent sadness" and reported sadness" featured in Montgomery-Åsberg Depression Rating Scale). We found no such pairs in our analysis. Next, roughly oppositely worded items such as "I am less shy or inhibited" and "I am more self-confident" were combined into one item within the HCL-32 questionnaire (reducing the number of its items from 32 to 31) in order to avoid further biasing the analysis by introducing two items measuring the same symptom. Subsequently, we examined each item pair across all of the scales, evaluating the overlap between every single item in every single scale (for all possible combinations). The content analysis resulted in a group of distinct symptoms categorized for each tool as (a) not featured in a scale, (b) featured as a part of a compound symptom, or (3) as a specific symptom.

According to the methodology of Fried (2017), items have been categorized as disparate items only if they clearly differ from each

other in terms of assessed symptom. Thus, items were categorized as equals with the same item content across scales as long as they were similarly worded, such as "I need less sleep" (HCL-32) and "... you got much less sleep than usual and found you didn't really miss it?" (MDQ) and "Sleep less and not feel tired" (MSQ). We differentiated between specific symptoms such as being "energetic" and "lack of energy and motivation," or between drinking "more coffee" and using other stimulants as it has been shown that specific symptoms should not be analyzed separately instead of gathering in compound items, for example, changes in energy levels, or taking "more stimulants" (Fried, 2017; Fried & Nesse, 2015). To cautiously evaluate whether items are disparate from each other, however, specific symptoms such as being more "physically active" in the MDQ and compound symptoms such as being "more active" in HCL-32 were considered to be overlapping, seeing that one is sufficient for fulfilling the other (Fried, 2017). Not recognizing specific and compound symptoms as overlapping would lead to an increase of heterogeneity and idiosyncrasy of the analyzed scales (Fried, 2017). Items have been categorized as idiosyncratic, when they were unique to a particular scale, for example, "feel one with nature" appeared only in the MSQ. Additionally, we counted how many items from each scale correspond to the DSM-5 criterion symptoms for a manic/hypomanic episode.

2.2 | Statistical analyses

Overlap of the content was estimated with the use of the Jaccard Index, a similarity coefficient that ranges from 0 (no overlap among scales) to 1 (complete overlap). The Jaccard index was calculated by $s/(u1 + u2 + s)$, where s is the number of items two questionnaires share, and $u1$ and $u2$ the number of items that are unique to each of the two scales (Fried, 2017). The criteria of the Jaccard Index power were very weak 0.00–0.19, weak 0.20–0.39, moderate 0.40–0.59, strong 0.60–0.79, and very strong 0.80–1.0 (as in Fried, 2017, based on Evans, 1996). We calculated the rate of specific (e.g., being more "physically active") versus compound symptoms (e.g., being "more active") per scale. Additionally, we counted the rate of idiosyncratic symptoms per scale, that is, symptoms that do not occur in other scales. In order to answer the question whether the length of the scale is associated with a higher overlap of its symptoms with the other scales measured in this paper, we measured the correlation between the mean Jaccard coefficient of each scale and the length of the scale for the number of specific symptoms captured and the adjusted for scale length (Fried, 2017). Analyses were conducted using R software with the use of the code supplied in (Fried, 2017).

3 | RESULTS

Analysis of 128 items across seven rating scales resulted in the extraction of 64 disparate (hypo)manic symptoms (Figure 1).

Symptoms are represented as a mean of three out of the seven analyzed tools (mode = 1, median = 2). From the group of 64 symptoms, 23 (36%) were present only in one single scale, whereas four symptoms appeared in all of the scales: elevated mood, irritability, talkativeness, and increased sexual interest. Three of them, elevated

mood, irritability, and talkativeness, were specifically captured by every instrument. Table 1 presents in how many of the scales the symptoms are listed.

Apart from elevated mood, irritability, and talkativeness, other DSM-5 core symptoms are represented accordingly: needing less sleep, race of thoughts and increased energy in six scales, increased engaging into new activities and distractibility in five scales, and risk-taking in two scales.

3.1 | Scale properties and overlap

Table 2 presents the degree of the symptoms' idiosyncrasy and specificity in each instrument, with the number of specific symptoms captured per tool and the adjusted scale length.

The largest number of idiosyncratic items (32%) is comprised in BSDS. MSQ and HCL-32 present 29% and 15% of the idiosyncratic symptoms. The other scales, YMRS, MDQ, BRMRS, and CARS-M, have no idiosyncratic symptoms. The largest number of all 64 disparate items is captured by the HCL-32 (51.6%). The lowest number is captured by the MDQ (35.9%).

The mean overlap among questionnaires estimated via the Jaccard index is 0.35. This implies a weak similarity between the scales (Evans, 1996). Table 3 presents the specific overlap among all of the analyzed tools.

The BSDS presents the lowest mean overlap with the other scales (0.29). The HCL-32, MSQ, and BSDS all present Jaccard coefficients below 0.40, which indicates a weak overlap with the other instruments. The YMRS, MDQ, BRMRS, and CARS-M present very similar average overlapping with the other scales (range 0.46–0.48; moderate overlap). The lowest overlap among the individual scales is between the BSDS and MSQ (0.20) indicating that these tools are composed of mostly different items. The highest overlap is observed between the CARSM and YMRS (0.65; strong overlap).

The correlation between the average Jaccard coefficient of each scale (the mean overlap of a scale with all of the others) and the length of the scale is 0.65 for the number of specific symptoms captured and 0.49 for the adjusted scale length. This indicates that longer scales overlap more with others and feature more representative content.

4 | DISCUSSION

In this study, we identified 64 disparate (hypo)manic symptoms in seven widely used rating scales. The mean overlap of the item content among these scales is weak and accounts to 0.35. Thirty-six percent of the symptoms appeared only in one scale. Only 6% of symptoms appeared in all of the instruments. The results of our analysis are similar to those obtained by Fried (2017) in which the author evaluated the overlap between 52 depression symptoms in seven common depression scales. Fried (2017) showed that 40% of all of the symptoms appeared only in a single scale and 12% across all instruments. We have similarly shown that using (hypo)manic scales as interchangeable measurements is problematic and may pose a threat to the generalizability and replicability of bipolar disorder research.
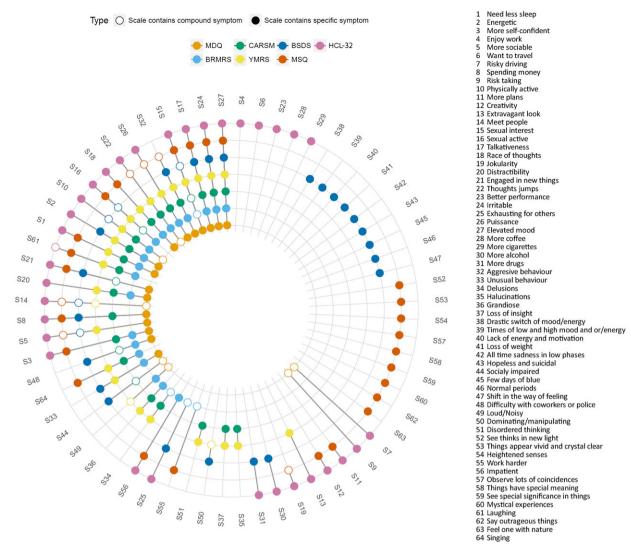
Type ○ Scale contains compound symptom ● Scale contains specific symptom

MDQ CARSM BSDS HCL-32
BRMRS YMRS MSQ

1 Need less sleep
2 Energetic
3 More self-confident
4 Enjoy work
5 More sociable
6 Want to travel
7 Risky driving
8 Spending money
9 Risk taking
10 Physically active
11 More plans
12 Creativity
13 Extravagant look
14 Meet people
15 Sexual interest
16 Sexual active
17 Talkativeness
18 Race of thoughts
19 Jokularity
20 Distractibility
21 Engaged in new things
22 Thoughts jumps
23 Better performance
24 Irritable
25 Exhausting for others
26 Puissance
27 Elevated mood
28 More coffee
29 More cigarettes
30 More alcohol
31 More drugs
32 Aggresive behaviour
33 Unusual behaviour
34 Delusions
35 Halucinations
36 Grandiose
37 Loss of insight
38 Drastic switch of mood/energy
39 Times of low and high mood and or/energy
40 Lack of energy and motivation
41 Loss of weight
42 All time sadness in low phases
43 Hopeless and suicidal
44 Socialy impaired
45 Few days of blue
46 Normal periods
47 Shift in the way of feeling
48 Difficulty with coworkers or police
49 Loud/Noisy
50 Dominating/manipulating
51 Disordered thinking
52 See thinks in new light
53 Things appear vivid and crystal clear
54 Heightened senses
55 Work harder
56 Impatient
57 Observe lots of coincidences
58 Things have special meaning
59 See special significance in things
60 Mystical experiences
61 Laughing
62 Say outrageous things
63 Feel one with nature
64 Singing

FIGURE 1 Co-occurrence of 64 (hypo)manic symptoms across seven (hypo)manic rating scales. Colored circles for a symptom indicating that a scale directly assesses that symptom, whereas empty circles indicate that a scale only measures a symptom indirectly. HCL-32: Hypomania Checklist 32; YMRS: Young Mania Rating Scale; MDQ: Mood Disorder Questionnaire; BSDS: Bipolar Spectrum Disorder Scale; BRMRS: The Bech-Rafaelsen Mania Rating Scale; CARS-M: Clinician-Administered Rating Scale For Mania; MSQ: Mood Swings Questionnaire

TABLE 1 Number of symptoms that appear across combinations of scales

Symptoms	Scales	%
23	1	36
13	2	20
6	3	9
3	4	5
7	5	11
8	6	13
4	7	6

To our best knowledge, this is the first study analyzing the content overlap of (hypo)manic rating scales. Previous studies concerning the tools for the assessment of bipolar disorder symptoms presented overviews (Miller et al., 2009; Rucci et al., 2013; Vieta, 2010) or cross-validations of the selected scales (Carta et al., 2006; Rucci et al., 2013; Rybakowski et al., 2012; Sasdelli et al., 2013). A solid

amount of studies compared the properties of HCL-32 and MDQ. A recent meta-analysis showed that the HCL-32 is significantly more accurate than the MDQ for the detection of BD II (Carvalho et al., 2015). It was shown that the HCL-32 presents a higher ability to identify the subgroup with hypomanic symptoms among major depression disorder patients than the MDQ (Rybakowski et al., 2012). It has been proposed that the differences between the two scales may rise from the different hypomanic symptoms listed in the two questionnaires. The MDQ includes a series of questions derived from the DSM-IV criteria of a manic episode (Hirschfeld et al., 2000), whereas symptoms of the HCL-32 are more numerous and pertain to various specific areas of everyday activity that may be changed during periods of hypomania (Rybakowski et al., 2012). Our results support this idea by showing that the HCL-32 captures the highest percent (51.6%) of all 64 disparate items among the seven analyzed scales, and the MDQ presents the lowest (35.9%). Additionally, the HCL-32 presents a higher number of idiosyncratic items than MDQ (5 vs. 0). The MDQ presents a higher overlap with other mania rating scales (YMRS,

TABLE 2 Idiosyncratic, specific, compound symptoms and percent of captured 64 disparate items per scale

	Symptoms captured (n)	Adjusted scale length (n)	Idiosyncratic items (n)	Idiosyncratic items (%)	Specific Items (%)	Compound items (%)	Scale captures x% of all 64 disparate items	Number of DSM-5 core mania symptoms
HCL-32	33	31	5	15	97	3	51.6	9
YMRS	26	9	0	0	88.5	11.5	40.6	7
MDQ	23	13	0	0	69.6	30.4	35.9	9
BSDS	28	18	9	32	82.1	17.9	43.8	5
BRMRS	24	19	0	0%	83.3	16.7	37.5	7
CARS-M	25	15	0	0	80	20	39.1	8
MSQ	31	27	9	29	80.6	19.4	48.4	6

Note. Symptoms captured present how many of 64 specific symptoms does the scale capture; adjusted scale length presents number of items per scale after combining similar items; idiosyncratic items rate the items that appear in no other scale, specific items present the rate of items that measure specific symptoms, compound items rate the items that measure compound symptoms. BRMRS: The Bech-Rafaelsen Mania Rating Scale; BSDS: Bipolar Spectrum Disorder Scale; CARS-M: Clinician-Administered Rating Scale for Mania; HCL-32: Hypomania Checklist 32; MDQ: Mood Disorder Questionnaire; MSQ: Mood Swings Questionnaire; YMRS: Young Mania Rating Scale.

TABLE 3 Overlap of item content of seven (hypo)manic scales

	HCL	YMRS	MDQ	BSDS	BRMRS	CARS-M	MSQ
HCL	1.00	0.37	0.47	0.30	0.39	0.38	0.42
YMRS	0.37	1.00	0.48	0.29	0.61	0.65	0.33
MDQ	0.47	0.48	1.00	0.38	0.52	0.55	0.35
BSDS	0.30	0.29	0.38	1.00	0.26	0.29	0.20
BRMRS	0.39	0.61	0.52	0.27	1.00	0.63	0.38
CARS-M	0.38	0.65	0.55	0.29	0.63	1.00	0.37
MSQ	0.42	0.33	0.35	0.20	0.37	0.37	1.00
Mean overlap	0.39	0.46	0.46	0.29	0.46	0.48	0.34

Note. The Jaccard Index ranges from 0 (no overlap) to 1 (total overlap). BRMRS: The Bech-Rafaelsen Mania Rating Scale; BSDS: Bipolar Spectrum Disorder Scale; CARS-M: Clinician-Administered Rating Scale for Mania; HCL-32: Hypomania Checklist 32; MDQ: Mood Disorder Questionnaire; MSQ: Mood Swings Questionnaire; YMRS: Young Mania Rating Scale.

BRMRS, CARSM, overlap: 0.48–0.55) than the HCL-32, MSQ, and BSDS. This result may be linked with the observations that the MDQ demonstrates a higher sensitivity to bipolar I disorder than in bipolar II/NOS disorder patients, and the patients' low rating of severity of mania explains almost half of all of the false negative results (Miller, Klugman, Berv, Rosenquist, & Nassir Ghaemi, 2004).

Our results indicate that the BSDS presented the highest number of idiosyncratic items (nine items, 32% of all items) and the lowest overlap with the other scales (0.29). This is the expected effect given the presence of questions concerning mood switches and depressive symptoms, not present in other scales. Also, this tool captured the lowest number of DSM-5 core mania symptoms (five out of nine). The BSDS was developed to identify milder forms of bipolar disorder. It has been shown that contrary to the MDQ, the sensitivity of the BSDS is similar for patients with BD I, BD II, and BD NOS/cyclothymia. However, its high negative predictive value and the low positive predictive value make this scale suitable for ruling out the diagnosis of bipolar disorder rather than ruling it in (Zimmerman, Galione, Chelminski, Young, & Ruggero, 2010).

Although the overlap between more hypomania-oriented scales (MDQ, HCL-32, BSDS, MSQ) varies from low (0.2 for MSQ vs. BSDS) to medium (0.47 for HCL-32 vs. MDQ), more mania-oriented scales (YMRS, BRMRS, and CARSM) present aligned, strong levels of overlap

(range: 0.63–0.65). Studies in which measures are based on mania-oriented scales are composed mostly of clinical trials or biomarker studies. Our results suggest that the YMRS, BRMRS, and CARMRS pose a relatively low risk of improper generalizability of mania scores across studies using these scales, due to the high overlap of the assessed symptoms. Slightly different, studies focused on screening for bipolar disorder or measuring bipolarity across different clinical or nonclinical populations are using hypomania-oriented scales, which present low or medium overlap and the highest rates of idiosyncratic items. Importantly, the difference in the overlap is also reflected by the coverage of DSM-5 core hypo(mania) symptoms. Whereas the HCL-32 and MDQ encompass all nine core symptoms, the BSDS and MSQ gather five and six, respectively.

Noteworthy, the MDQ, HCL-32, BSDS, and MSQ differ significantly in their factor structures. It has been shown that the HCL-32 consists of two factors: elated mood/energy, reflecting a more positive aspect of hypomania (being active, elated, self-confident, and cognitively enhanced); and a risk-taking/irritability factor (being impulsive, careless, and more substance use; Angst et al., 2010). The BSDS is composed of "Irritable and hyper-energetic" and "depressed and lack of energy" dimensions (Chu et al., 2010). The MSQ is distinguished by the highest number of factors: mood elevation, disinhibition, irritation, and mysticism (Parker, Hadzi-Pavlovic, & Tully, 2006). The latter

one is associated with the highest number of MSQ idiosyncratic items such as "see things in new light," "observing lots of coincidences," "heightened senses." The most recent study explicating the structure and relations of the MDQ, divided it into "Positive Activation" dimension, defined by the items assessing increased energy/activity, grandiosity, and decreased need for sleep, and "Negative Activation" associated with impulsivity, distractibility, racing thoughts, and risk-taking. Importantly, the mentioned distinct factors are related to different clinical and personality profiles (Stanton & Watson, 2017). Therefore, due to a low overlap of items and the differences of factors measured by the mentioned scales, we recommend against the use of their total scores as interchangeable measures of (hypo)manic symptoms.

Future studies assessing (hypo)manic symptoms could take into account the implications arising from our results. First, the weak mean overlap across analyzed scales obtained in our paper may indicate that the results of various studies using these scales may differ depending on the tool selected to assess (hypo)manic symptoms. Thus, we recommend using more than one scale. Such an attempt would allow to better understand the role of particular (hypo)manic symptoms in different research areas. It may also protect studies' results from being limited to a particular scale and its factors structure. Until now, there has been a limited number of studies comparing more than two (hypo)manic scales. Second, systematic reviews and research studies should be performed to identify the most important (hypo)manic symptoms. As an example, Scott et al. (2017) conducted a systematic review to determine whether activation is a distinct dimension of BD. Although evidence supports increased activation is a criterion A symptom of BD, the review suggests that despite the acknowledgement of this criterion for more than a century, this critical construct of (hypo)mania is understudied and should be the topic of more systematic high-quality research.

The limitation of this study is the choice of the analyzed scales and their number. Similar to the study of Fried (2017), we selected seven scales based on their frequency in the literature, inclusion in recent reviews, and citation count. A greater number of scales would probably increase the heterogeneity and unlikely decrease the low degree of overlap, and including different scales would unlikely change the results towards less heterogeneity or more overlap (Fried, 2017). The second limitation is the lack of an objective method of carrying out the comparisons of items across the scales because there is no method to evaluate whether two similarly worded symptoms are meant to measure the same problem or not (Fried, 2017). To counter this limitation, the authors used the conservative approach recommended by Fried, 2017: erring whenever it was possible, on the side of caution; considering symptoms rather too similar than too different. The third limitation is the presence of the BSDS questions concerning mood switches and depressive symptoms in the analysis. We decided to keep these items in the analysis because we aimed to evaluate the whole rating scales instead of selected symptoms.

To our best knowledge, we have performed the first analysis of content overlap between common (hypo)mania rating scales. We identified 64 disparate (hypo)manic symptoms in seven assessment tools and determined that the mean overlap of the item content among the scales is weak. Furthermore, 36% of the symptoms appeared only on one scale, with only 6% of symptoms appearing on all of the instruments. We have shown that using (hypo)manic scales as interchangeable measurements may be problematic, especially in the studies using the HCL-32, MDQ, BSDS, and MSQ scales for BD screening and bipolarity assessment in clinical and nonclinical populations. Given the low overlap between both depression, as well as (hypo)manic scales, this may pose a threat to the generalizability and replicability of BD research.

DECLARATION OF INTEREST STATEMENT

None to declare.

CONTRIBUTORS

A. A. C. designed the study, managed literature search, obtained data, conducted statistical analysis, and wrote the first draft and the revisions of the manuscript. M. S. designed the study and contributed to the revisions of the manuscript. D. D. designed the study and contributed to the revisions of the manuscript. J. K. R. designed the study, managed literature searches, and wrote the final version of the manuscript. All authors have approved the final version of the manuscript.

ORCID

Adrian A. Chrobak http://orcid.org/0000-0002-2107-9980

REFERENCES

Akiskal, H. S. (2007). The emergence of the bipolar spectrum: Validation along clinical-epidemiologic and familial-genetic lines. *Psychopharmacology Bulletin*, *40*, 99–115.

Angst, J., Ajdacic-Gross, V., & Rössler, W. (2015). Classification of mood disorders. *Psychiatria Polska*, *49*(4), 663–671. https://doi.org/10.12740/PP/58259

Angst, J., Cui, L., & Swendsen, J. (2010). Major depressive disorder with sub-threshold bipolarity in the National Comorbidity Survey Replication. *The American Journal of Psyhiatry*, *167*(10), 1194–1201. https://doi.org/10.1176/appi.ajp.2010.09071011.Major

Angst, J., Meyer, T. D., Adolfsson, R., Skeppar, P., Carta, M., Benazzi, F., … Gamma, A. (2010). Hypomania: A transcultural perspective. *World Psychiatry*, *9*(1), 41–49. https://doi.org/10.1002/j.2051-5545.2010.tb00268.x

Antosik-Wójcińska, A., Stefanowski, B., & Święcicki, Ł. (2015). Efficacy and safety of antidepressant's use in the treatment of depressive episodes in bipolar disorder—Review of research. *Psychiatria Polska*, *49*(6), 1223–1239. https://doi.org/10.12740/PP/37914

Bauer, M. S., Vojta, C., Kinosian, B., Altshuler, L., & Glick, H. (2000). The Internal State Scale: Replication of its discriminating abilities in a multi-site, public sector sample. *Bipolar Disorders*, 2(4), 340–346. https://doi.org/10.1034/j.1399-5618.2000.020409.x

Bech, P. (2002). The Bech-Rafaelsen Mania Scale in clinical trials of therapies for bipolar disorder: A 20-year review of its use as an outcome measure. *CNS Drugs*, 16(1), 47–63. doi:160104 [pii]

Carta, M. G., Hardoy, M. C., Cadeddu, M., Murru, A., Campus, A., Morosini, P. L., ... Angst, J. (2006). The accuracy of the Italian version of the Hypomania Checklist (HCL-32) for the screening of bipolar disorders and comparison with the Mood Disorder Questionnaire (MDQ) in a clinical sample. *Clinical Practice and Epidemiology in Mental Health: CP & EMH*, 2, 2. https://doi.org/10.1186/1745-0179-2-2

Carvalho, A. F., Takwoingi, Y., Sales, P. M. G., Soczynska, J. K., Köhler, C. A., Freitas, T. H., ... Vieta, E. (2015). Screening for bipolar spectrum disorders: A comprehensive meta-analysis of accuracy studies. *Journal of Affective Disorders*, 172, 337–346. https://doi.org/10.1016/j.jad.2014.10.024

Chu, H., Lin, C.-J., Chiang, K.-J., Chen, C.-H., Lu, R.-B., & Chou, K.-R. (2010). Psychometric properties of the Chinese version of the Bipolar Spectrum Diagnostic Scale. *Journal of Clinical Nursing*, 19(19–20), 2787–2794. https://doi.org/10.1111/j.1365-2702.2010.03390.x

Dell'Osso, L., Armani, A., Rucci, P., Frank, E., Fagiolini, A., Corretti, G., ... Cassano, G. B. (2002). Measuring mood spectrum: Comparison of interview (SCI-MOODS) and self-report (MOODS-SR) instruments. *Comprehensive Psychiatry*, 43(1), 69–73. https://doi.org/10.1053/comp.2002.29852

Dudek, D., Siwek, M., Jaeschke, R., Drozdowicz, K., Styczeń, K., Arciszewska, A., ... Rybakowski, J. K. (2015). A web-based study of bipolarity and impulsivity in athletes engaging in extreme and high-risk sports. *Acta Neuropsychiatrica*, 28, 1–5. https://doi.org/10.1017/neu.2015.44

Evans, J. (1996). *Straightforward statistics for the behavioral sciences*. Pacific Grove, CA: Brooks/Cole publishing.

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*, 208(October), 191–197. https://doi.org/10.1016/j.jad.2016.10.019

Fried, E. I., & Nesse, R. M. (2015). Depression sum-scores don't add up: Why analyzing specific depression symptoms is essential. *BMC Medicine*, 13(1), 72. https://doi.org/10.1186/s12916-015-0325-4

Ghaemi, S. N., Ko, J. Y., & Goodwin, F. K. (2001). The bipolar spectrum and the antidepressant view of the world. *Journal of Psychiatric Practice*, 7(5), 287–297. https://doi.org/10.1097/00131746-200109000-00002

Hirschfeld, R. M. A., Williams, J. B. W., Spitzer, R. L., Calabrese, J. R., Flynn, L., Keck, P. E., ... Zajecka, J. (2000). Development and validation of a screening instrument for bipolar spectrum disorder: The Mood Disorder Questionnaire. *American Journal of Psychiatry*, 157(11), 1873–1875. https://doi.org/10.1176/appi.ajp.157.11.1873

Judd, L. L., & Akiskal, H. S. (2003). The prevalence and disability of bipolar spectrum disorders in the US population: Re-analysis of the ECA database taking into account subthreshold cases. *Journal of Affective Disorders*, 73(1–2), 123–131.

Klimkiewicz, A., Klimkiewicz, J., Jakubczyk, A., Kieres-Salomoński, I., & Wojnar, M. (2015a). Comorbidity of alcohol dependence with other psychiatric disorders, part I—Epidemiology of dual diagnosis. *Psychiatria Polska*, 49(2), 265–275. https://doi.org/10.12740/PP/25704

Klimkiewicz, A., Klimkiewicz, J., Jakubczyk, A., Kieres-Salomoński, I., & Wojnar, M. (2015b). Comorbidity of alcohol dependence with other psychiatric disorders, part II—Pathogenesis and treatment. *Psychiatria Polska*, 49(2), 277–294. https://doi.org/10.12740/PP/26071

Lee, D., Cha, B., Park, C. S., Kim, B. J., Lee, C. S., & Lee, S. (2013). Usefulness of the combined application of the Mood Disorder Questionnaire and Bipolar Spectrum Diagnostic Scale in screening for bipolar disorder. *Comprehensive Psychiatry*, 54(4), 334–340. https://doi.org/10.1016/j.comppsych.2012.10.002

Łojko, D., Dudek, D., Angst, J., Siwek, M., Michalak, M., & Rybakowski, J. K. (2016). The 33-item Hypomania Checklist (HCL-33)—A study of the consistency between self- and external assessments in Polish bipolar patients. *Psychiatria Polska*, 50(6), 1085–1092. https://doi.org/10.12740/PP/66358

Merikangas, K. R., Jin, R., He, J.-P., Kessler, R. C., Lee, S., Sampson, N. A., ... Zarkov, Z. (2011). Prevalence and correlates of bipolar spectrum disorder in the World Mental Health Survey Initiative. *Archives of General Psychiatry*, 68(3), 241–251. https://doi.org/10.1001/archgenpsychiatry.2011.12

Miller, C. J., Johnson, S. L., & Eisner, L. (2009). Assessment tools for bipolar disorder. *Clin Psychol (New York)*, 16(2), 188–201. https://doi.org/10.1111/j.1468-2850.2009.01158.x.Assessment

Miller, C. J., Klugman, J., Berv, D. A., Rosenquist, K. J., & Nassir Ghaemi, S. (2004). Sensitivity and specificity of the Mood Disorder Questionnaire for detecting bipolar disorder. *Journal of Affective Disorders*, 81(2), 167–171. https://doi.org/10.1016/S0165-0327(03)00156-3

Parker, G., & Fletcher, K. (2013). The "real world" utility of a web-based bipolar disorder screening measure: A replication study. *Journal of Affective Disorders*, 150(2), 276–283. https://doi.org/10.1016/j.jad.2013.04.006

Parker, G., Fletcher, K., Blanch, B., & Greenfield, L. (2012). Take-up and profile of individuals accessing a web-based bipolar self-test screening measure. *Journal of Affective Disorders*, 138(1–2), 117–122. https://doi.org/10.1016/j.jad.2011.12.013

Parker, G., Fletcher, K., McCraw, S., & Hong, M. (2014). The hypomanic personality scale: A measure of personality and/or bipolar symptoms? *Psychiatry Research*, 220(1–2), 654–658. https://doi.org/10.1016/j.psychres.2014.07.040

Parker, G., Graham, R., Hadzi-Pavlovic, D., Fletcher, K., Hong, M., & Futeran, S. (2012). Further examination of the utility and comparative properties of the MSQ and MDQ bipolar screening measures. *Journal of Affective Disorders*, 138(1–2), 104–109. https://doi.org/10.1016/j.jad.2011.12.016

Parker, G., Graham, R., Rees, A. M., Futeran, S., & Friend, P. (2012). A diagnostic profile of those who return a false positive assignment on bipolar screening measures. *Journal of Affective Disorders*, 141(1), 34–39. https://doi.org/10.1016/j.jad.2012.02.026

Parker, G., Graham, R., Synnott, H., & Anderson, J. (2014). Is the DSM-5 duration criterion valid for the definition of hypomania? *Journal of Affective Disorders*, 156, 87–91. https://doi.org/10.1016/j.jad.2013.11.020

Parker, G., & Graham, R. K. (2015). An evaluation of the DSM-5 rules defining mania and hypomania with identical symptom criteria. *Journal of Affective Disorders*, 170, 91–94. https://doi.org/10.1016/j.jad.2014.08.051

Parker, G., Hadzi-Pavlovic, D., & Tully, L. (2006). Distinguishing bipolar and unipolar disorders: An isomer model. *Journal of Affective Disorders*, 96(1–2), 67–73. https://doi.org/10.1016/j.jad.2006.05.014

Rajewska-Rager, A., Sibilski, P., & Lepczyńska, N. (2015). Risk factors for suicide among children and youths with bipolar spectrum and early bipolar disorder. *Psychiatria Polska*, 49(3), 477–488. https://doi.org/10.12740/PP/29415

Rucci, P., Calugi, S., Miniati, M., & Fagiolini, A. (2013). A review of self-report and interview-based instruments to assess mania and hypomania symptoms. *Journal of Psychopathology*, 19(2), 143–159.

Rybakowski, J. K., Dudek, D., Pawlowski, T., Lojko, D., Siwek, M., & Kiejna, A. (2012). Use of the Hypomania Checklist-32 and the Mood Disorder Questionnaire for detecting bipolarity in 1,051 patients with major depressive disorder. *European Psychiatry*, 27(8), 577–581. https://doi.org/10.1016/j.eurpsy.2010.12.001

Sasdelli, A., Lia, L., Luciano, C. C., Nespeca, C., Berardi, D., & Menchetti, M. (2013). Screening for bipolar disorder symptoms in depressed primary care attenders: Comparison between Mood Disorder Questionnaire and Hypomania Checklist (HCL-32). *Psychiatry Journal*, 2013, 548349–548349. https://doi.org/10.1155/2013/548349

Scott, J., Murray, G., Henry, C., Morken, G., Scott, E., Angst, J., ... Hickie, I. B. (2017). Activation in bipolar disorders a systematic review. *JAMA Psychiatry*, *74*(2), 189–196. https://doi.org/10.1001/jamapsychiatry.2016.3459

Siwek, M., Dudek, D., Drozdowicz, K., Jaeschke, R., Styczen, K., Arciszewska, A., ... Rybakowski, J. K. (2015). Temperamental dimensions of the TEMPS-A in male and female subjects engaging in extreme or/and high risk sports. *Journal of Affective Disorders*, *170*, 66–70. https://doi.org/10.1016/j.jad.2014.08.036

Siwek, M., Dudek, D., Jaeschke, R., Dembińska-Kieć, A., Witkowski, L., Arciszewska, A., ... Rybakowski, J. (2015). Bipolar spectrum features in obese individuals. *Psychiatria Polska*, *49*(5), 993–1004. https://doi.org/10.12740/PP/OnlineFirst/32937

Siwek, M., Dudek, D., Rybakowski, J., Łojko, D., Pawłowski, T., & Kiejna, A. (2009). Kwestionariusz Zaburzeń Nastroju - Charakterystyka i zastosowanie. *Psychiatria Polska*, *43*(3), 287–299.

Stanton, K., & Watson, D. (2017). Explicating the structure and relations of the Mood Disorder Questionnaire: Implications for screening for bipolar and related disorders. *Journal of Affective Disorders*, *220*, 72–78. https://doi.org/10.1016/j.jad.2017.05.046

Vieta, E. (2010). *Guide to assessment scales in bipolar disorder* (Second ed.). London: Current Medicine Group Ltd.

Vieta, E., Berk, M., Schulze, T. G., Carvalho, A. F., Suppes, T., Calabrese, J. R., ... Grande, I. (2018). Bipolar disorders. *Nature Reviews Disease Primers*, *4*(March), 18008. https://doi.org/10.1038/nrdp.2018.8

Young, R. C., Biggs, J. T., Ziegler, V. E., & Meyer, D. A. (1979). A rating scale for mania: Reliability, Validity and Sensitivity. *Br. J. Psychiatry*, *133*, 429–435.

Zimmerman, M., Galione, J. N., Chelminski, I., Young, D., & Ruggero, C. J. (2010). Performance of the Bipolar Spectrum Diagnostic Scale in psychiatric outpatients. *Bipolar Disorders*, *12*(5), 528–538. https://doi.org/10.1111/j.1399-5618.2010.00840.x